

Google si auto-concede di usare i dati di internet per le sue IA

In una mossa che denota una certa eleganza tattica, **Google ha modificato le proprie policy** per notificare al mondo intero ciò che molti sospettavano già, ovvero che la Big Tech abbia intenzione di **esfiltrare** ogni angolo della Rete al fine di accumulare una mole incalcolabile di dati che poi verranno riversati nell'**addestramento di nuovi prodotti** legati all'**intelligenza artificiale**. Fino a pochi giorni fa, l'azienda riconosceva la raccolta dei dati, ma dichiarava che l'obiettivo perseguito fosse quello di perfezionare gli strumenti linguistici di traduzione.

“Google usa le informazioni per migliorare i suoi servizi e **sviluppare nuovi prodotti**, funzionalità e tecnologie che beneficiano i nostri utenti e il pubblico”, riporta l'[alterazione](#) del testo datata primo luglio. “Per esempio, usiamo le informazioni disponibili pubblicamente per aiutare l'addestramento dei modelli AI di Google e per costruire prodotti e funzionalità quali Google Translate, Bard e le potenzialità della Cloud AI”.

Se la raccolta massiva e incondizionata dei dati - il cosiddetto **data scraping** - è da sempre motivo di preoccupazione per coloro che si interessano a diritti e privacy, il fatto che una simile pratica sia utilizzata esplicitamente per sviluppare prodotti di consumo non manca di sollevare animate discussioni. Già ora, attivisti e avvocati stanno cercando di valutare se l'ultimo passo della Big Tech possa in qualche modo essere comparato alle manovre di ClearviewAI, azienda che con l'esfiltrazione delle informazioni aveva ricavato un sistema di sorveglianza basato sui *social* che è stato poi considerato illegittimo da molteplici nazioni.

Questa apparentemente minuscola ridefinizione delle policy aziendali verrà quasi certamente impugnata dagli attivisti per esercitare pressioni nei confronti dei legislatori, soprattutto quelli europei, i quali saranno chiamati a disegnare **confini più chiari** per quanto concerne i metodi di raccolta dei dati. La legalità di questa forma di “pesca a strascico” della sfera digitale è spesso messa in dubbio e la sua deriva commerciale è motivo di accesi dibattiti [legislativi ed economici](#). Se fino a non troppo tempo fa l'attenzione era incentrata sull'accessibilità dei dati, a essere ora urgente è la valutazione di **come queste informazioni possano essere utilizzate**, nonché sull'impatto che le stesse possono sviluppare sul tessuto sociale.

Se ormai è assodato che tutto ciò che finisce pubblicamente *online* sarà alla portata di tutti, si sta progressivamente diffondendo la consapevolezza che i contenuti presenti in chiaro su internet sono reclamati da alcune aziende *tech* come risorse **di loro stessa proprietà**, senza che ci sia troppa cura nel tutelare *privacy* e diritti d'autore. Aziende quali [Twitter](#) e [Reddit](#) hanno fatto riferimento proprio all'esfiltrazione delle informazioni da parte dei sistemi di addestramento delle IA per giustificare la scelta - altrettanto controversa - di sospendere il libero accesso ai loro servizi, un fenomeno reattivo che a sua volta prelude alla

Google si auto-concede di usare i dati di internet per le sue IA

possibile ridefinizione di come il *web* sia vissuto da aziende e utenti.

Nonostante la forza esercitata dalle [lobby delle Big Tech](#), è possibile che la rivisitazione delle policy di Google finisca con il diventare argomento di discussione durante le **contrattazioni** degli europarlamentari che dovranno ragionare sull'applicazione dell'[AI Act](#). Secondo il pacchetto di norme approvato dal Parlamento UE, le aziende operanti con le intelligenze artificiali dovranno infatti **esplicitare tutte le fonti protette da copyright** utilizzate per addestrare le loro macchine, un traguardo difficilmente raggiungibile quando le informazioni vengono raccolte massivamente e con approccio indiscriminato. Non a caso Sam Altman, CEO di OpenAI, azienda legata a Microsoft, ha dato a intendere che sia pronto ad abbandonare l'Europa, qualora le leggi di tutela dei dati si dimostrassero troppo rigide.

[di Walter Ferri]